Measuring Task-Agnostic Training Data Influence Across Language Model Pretraining

Yuto Nishida[1,2] Hirokazu Kiyomaru[2] Yusuke Oda[2] Takashi Kodama[2] Chaoran Liu[2]
Daisuke Kawahara[2,3] Yusuke Miyao[2,4] Max Müller-Eberstein[4,5] Masaru Isonuma[2,6]

[1]Nara Institute of Science and Technology [2]NII LLMC
[3]Waseda University [4]The University of Tokyo [5]IT University of Copenhagen [6]Tohoku University

1. Key Takeaways

- **Task-agnostic training data influence** measures progress towards the final pretrained parameters, not just individual downstream tasks.
- **Influential data shift over training: literature-related examples early, STEM examples later.**

2. Introduction

Background: Training data influence

- Which examples shape a model, and when?

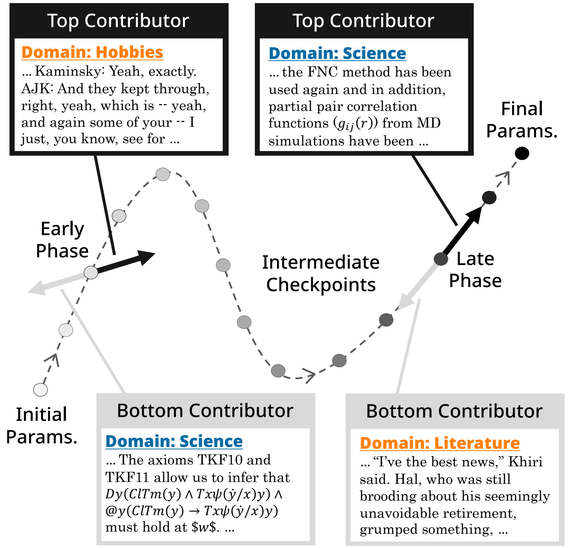
Challenge: Assessing pretraining influence

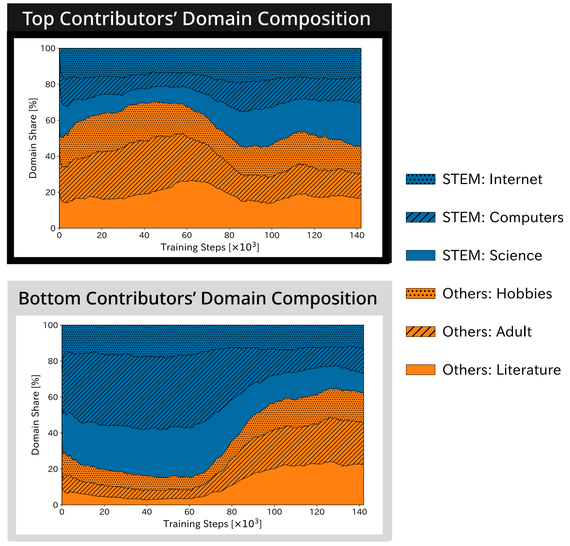
- Most methods measure influence on a specific task, capturing only part of training dynamics.

Our approach: A task-agnostic measure

- **Use the final parameters as a common reference for model capabilities; measure update alignment toward the final model.**

3. Methodology & Main Results





Intuition

- **Influential examples move the model toward its final pretrained parameters.**

How we measure it

- We quantify how much each example's update reduces the squared L2 distance to the final parameters.
  - **Larger reduction → higher influence.**

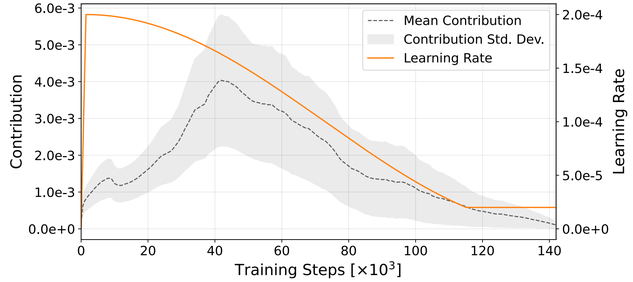
Experiments & Domain shift

- Main plot: Pythia-1.4B-Deduped.
- Estimate contribution on actual training-stream examples from checkpoints.
- Classify examples into 26 domains; compare the domain mix of the top/bottom 5%.
- **Literature-related data are influential early; STEM data become influential later.**

4. Analysis

Contribution Dynamics

- Mean contribution peaks in the middle stage, then declines toward the end.



Text Difficulty (PPL) vs. Contribution

- Use Pythia-12B PPL as a difficulty signal.
- 1,000 examples per training interval; split into 5 PPL bins.
- **Middle stage: higher-PPL → larger share of contribution.**
- Late stage: flatter across PPL bins.

